AD-A054 824 YALE UNIV NEW HAVEN CT DEPT OF PSYCHOLOGY F/G 5/10

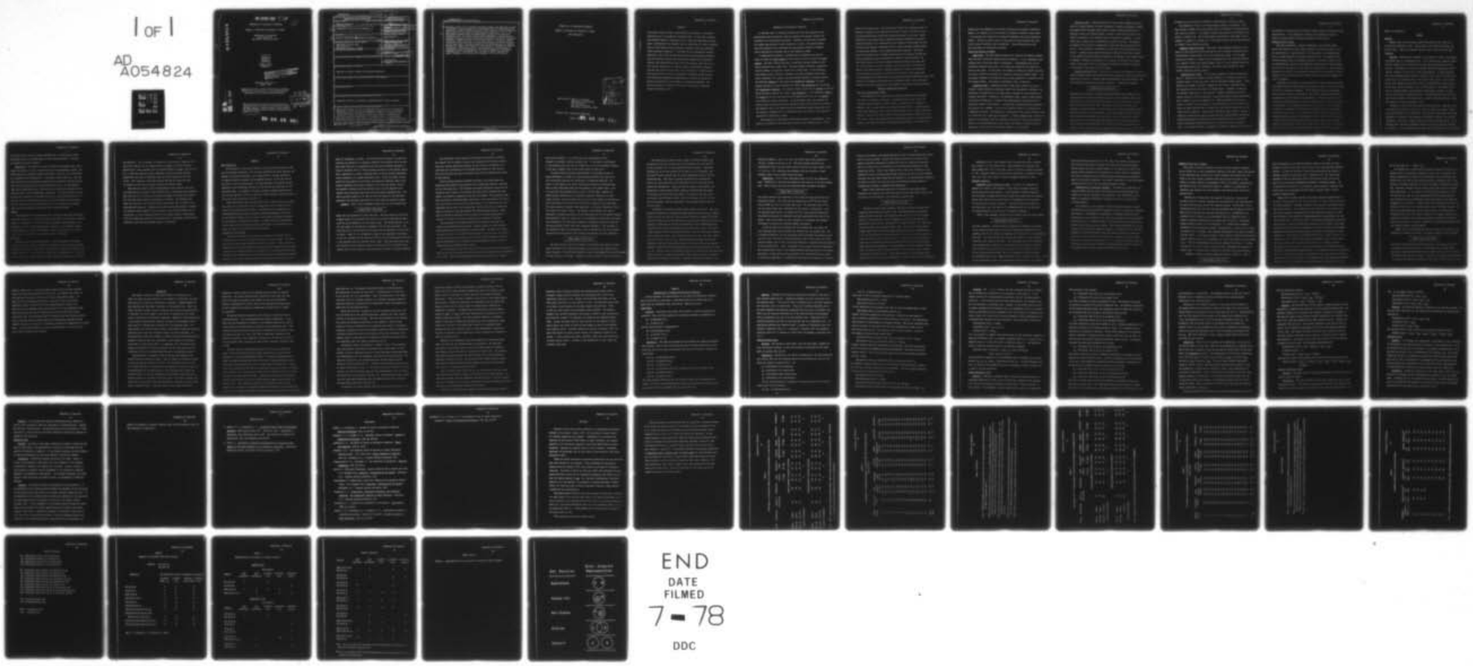
COMPONENTS OF SYLLOGISTIC REASONING.(U)

APR 78 R J STERNBERG, M E TURNER N00014-78-C-0025

UNCLASSIFIED RR-6-78 NL

1 OF 1

AD A054824

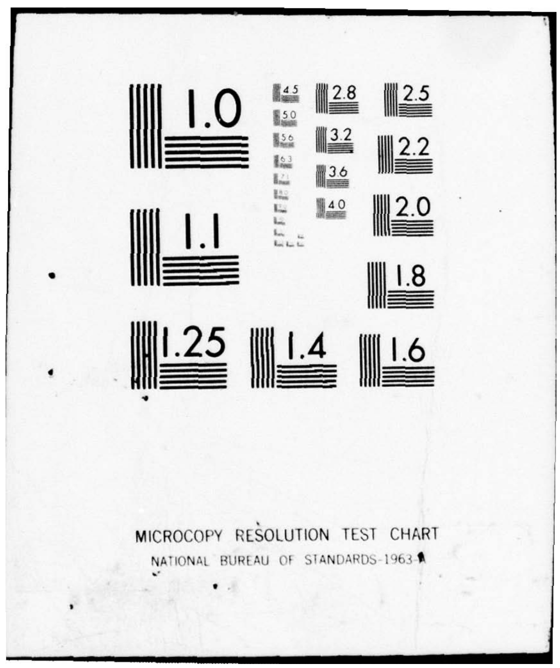


END

DATE FILMED

7-78

DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Components of Syllogistic Reasoning

Robert J. Sternberg and Margaret E. Turner

Department of Psychology
Yale University
New Haven, Connecticut 06520

Technical Report No. 6
April, 1978

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N0001478C0025, Contract Authority Identification Number NR 150-412.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14 RR-6-78

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report N 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 Components of Syllogistic Reasoning.		5. ~~TYPE OF REPORT & PERIOD COVERED~~ ~~Periodic~~ Technical Report. no. 6, 1 Jan 78 - 30 Mar 78,
		6. ~~PERFORMING ORG. REPORT NUMBER~~ Research Report No. 6-78
7. AUTHOR(s) 10 Robert J./Sternberg [redacted] Margaret E./Turner		8. CONTRACT OR GRANT NUMBER(s) N0001478C0025 15 N00014-78-C-0025
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Yale University New Haven, Connecticut 06520		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N 16 RR-042-04; 17 RR-042-04-01; NR 150-412
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 11 1 Apr 78
		13. NUMBER OF PAGES 57 12 60p.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

A049 888

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

reasoning, deduction, component, transitive-chain theory, syllogism

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The present research sought to understand the components of syllogistic reasoning that are used in a syllogistic evaluation task, for example, "Some humbugs study syllogistic reasoning; some Yale professors study syllogistic reasoning; therefore, some Yale professors are humbugs" (is this conclusion definitely true, possibly true, or never true?). The research had two major goals. The first was to compare one particular model of syllogistic reasoning, the transitive-chain model (Guyote ~~&~~ and (over)

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

482 628 Gul

Sternberg, ~~Note 1~~), to plausible alternative models that have been proposed in the past. Recent comparisons of the models using a response-selection task have provided convincing evidence of the superiority of the transitive-chain model for this particular task (Guyote & Sternberg, Note 1), and the present research seeks to extend ~~these~~ these findings to the response-evaluation task. The second goal ~~of the research~~ was to separate experimentally the premise encoding and premise combination stages of syllogistic reasoning, thereby enabling (a) more direct tests of the various models' assumptions about each stage than has been possible in previous research, and (b) more direct inferences regarding the representations of relations between the subject and predicate of the premises as encoded and combined by subjects. This second goal was accomplished by a modified form of componential analysis (Sternberg, 1977, 1978), whereby an information-processing task is decomposed into a series of nested subtasks that permit isolation of the elementary components of task performance.

3

Components of Syllogistic Reasoning

Robert J. Sternberg and Margaret E. Turner

Yale University

Send proofs to Robert J. Sternberg
Department of Psychology
Yale University
Box 11A Yale Station
New Haven, Connecticut 06520

Running head: Syllogistic Reasoning

April, 1978

Abstract

The present research sought to understand the components of syllogistic reasoning used in a syllogistic evaluation task. In this task, subjects must indicate whether a conclusion such as "Some Yale professors are humbugs" is definitely true, possibly true, or never true of a set of premises such as "Some humbugs study syllogistic reasoning; some Yale professors study syllogistic reasoning." A modified form of componential analysis (Sternberg, 1977, 1978) was used to decompose the syllogistic evaluation task with abstract content into encoding and encoding plus combination subtasks. The response-choice data from these subtasks were used to provide (a) more direct tests of various models of syllogistic reasoning than has been possible in the past, and in particular, of their assumptions about sources of error in syllogistic reasoning; and (b) more direct inferences regarding the representation of relations between the subject and predicate of the premises as encoded and combined. The results supported a transitive-chain model of syllogistic reasoning (Guyote & Sternberg, Note 1).

Components of Syllogistic Reasoning

Of the many types of reasoning problems that have been studied by both psychologists and philosophers, none has induced more research than has the categorical (or Aristotelian) syllogism. Certainly, no other type of reasoning problem has retained so much interest for so long. The ancient Greeks were avid students of the syllogism, and the problem continues to generate active theoretical controversy even in the research of today.

A categorical syllogism is a problem with two premises, the first of which is called the major premise and the second of which is called the minor premise. The major premise describes a quantified relation between a syllogistic predicate and a middle term. The minor premise describes a quantified relation between a syllogistic subject and a middle term. In the syllogism, "Some humbugs study syllogistic reasoning; some Yale professors study syllogistic reasoning," for example, the major premise describes a relation between the predicate humbugs and the middle term syllogistic reasoning; the minor premise describes a relation between the subject Yale professors and the middle term syllogistic reasoning. The subject's task is either to (a) produce a logically valid conclusion relating the subject, Yale professors, to the predicate, humbugs, (b) select one of four conclusions as the logically valid one, or (c) evaluate the validity of a given conclusion. In the evaluation task, for example, the subject might be presented with the conclusion "Some Yale professors are humbugs." The subject's task is to decide whether the conclusion follows deductively from the premises, irrespective of its real-world truth or falsity. This particular conclusion is deductively invalid.

The research to be described in this article sought to understand the components of syllogistic reasoning that are used in the evaluation task. The research had two major goals. The first was to compare one particular model of syllogistic reasoning, the transitive-chain model (Guyote & Sternberg, Note 1), to plausible alternative models that have been proposed in the past. Recent comparisons of the models using a response selection paradigm have provided convincing evidence of the superiority of the transitive-chain model for this particular task (Guyote & Sternberg, Note 1), and the present research seeks to extend these findings to the response-evaluation paradigm. The second goal of the research was to separate experimentally the premise encoding and premise combination stages of syllogistic reasoning, thereby enabling (a) more direct tests of the various models' assumptions about each stage than has been possible in previous research, and (b) more direct inferences regarding the representations of relations between the subject and predicate of the premises as encoded and combined by the subjects. This second goal was accomplished by a modified form of componential analysis (Sternberg, 1977, 1978), whereby an information-processing task is decomposed into a series of nested subtasks that permits isolation of the elementary components of task performance.

Models of Syllogistic Reasoning

Bases for Distinguishing Models

The experiment to be described later permitted comparative tests among six models of syllogistic reasoning: a baseline "ideal-subject" model, an atmosphere model (Woodworth & Sells, 1935), a conversion model (Chapman & Chapman, 1959), a random-combination model (Erickson, 1974), a complete-combination model (Erickson, 1974), and a transitive-chain model (Guyote & Sternberg, Note 1).[1] There are two primary bases for distinguishing the predictions of the six models. First, the models differ in whether they predict that subjects make errors in (a) neither encoding nor combination of premises (ideal model), (b) encoding but not combination of premises (conversion and complete-combination models), (c) combination but not encoding of premises (transitive-chain model), or (d) encoding and combination of premises (atmosphere and random-combination models). Second, the models differ in the kinds of errors subjects are theorized to make in encoding and/or combination. These differences are what give each model its distinctive character.

<u>Verbal Descriptions of Models</u>

<u>Ideal model</u>. The ideal-subject model may be viewed as the baseline against which the other alternative models should be compared. It is a competence model assuming that no errors are made in performance. A performance model can be considered minimally viable only if its empirical predictions are superior to those of this model. In this model, the subject is viewed as an expert logician. The subject encodes all premises in a logically correct and complete manner. Then the subject combines the information that has been encoded in a logically correct and complete manner. The subject's performance, then, is flawless.

<u>Atmosphere model</u>. Woodworth and Sells (1935) took a rather dim view of subjects' logical abilities, proposing a model that was almost as extreme in its postulation of alogicality as the ideal model is extreme in its postulation of complete logicality. According to the atmosphere model, subjects always encode the polarity (affirmative or negative) and quantification (universal [all] or particular [some]) of each premise. In combining information from the premises, the presence of a negative in one or both premises leads to a preference for a negative conclusion. The presence of a particular in one or both premises leads to a preference for a particular conclusion. If both a negative and a particular appear in the premises, then the preferred conclusion is a particular negative. Note that in the limiting case of this model, the subject is utterly alogical, responding only on the basis of surface features of the premises.

Conversion model. Whereas Woodworth and Sells (1935) suggested that subjects are largely alogical in their combination of premise information, Chapman and Chapman (1959) suggested that subjects combine premise information in a logically correct and complete manner. Subjects' errors derive from the conversion of premises in the encoding stage of syllogistic reasoning. In other words, both a premise and its converse are assumed to be true. For example, the premise "All *A* are *B*" is interpreted to mean that "All *A* are *B* and all *B* are *A*." "Some *A* are *B*" is interpreted to mean that "Some *A* are *B* and some *B* are *A*." A problem arises because although the meaning of the latter premise is not changed when this premise is converted, the meaning of the former premise is changed. Similarly, the meaning of the premise "No *A* are *B*" (or, equivalently, "All *A* are not *B*") is not changed by conversion, whereas the meaning of the premise "Some *A* are not *B*" is changed.

Random-combination model. According to this model, as proposed by Erickson (1974), subjects encode only certain single representations of the set relations between the two terms of a premise; other representations are never encoded. The five possible set relations are shown in Figure 1. Consider, for example,

---

Insert Figure 1 about here

---

the premise "All *A* are *B*." Erickson has suggested that with probability .75, subjects encode the relation between *A* and *B* as one of equivalent sets; with probability .25, subjects encode the relation as one of subset (*A*) - set (*B*). Other set relations are not used. Note that although both relations are correct, only one is encoded. Note also that both of the two possible correct relations are used at least sometimes. This is not always the case, however. In representing the premise "Some *A* are *B*," subjects are assumed to encode the set (*A*) - subset (*B*) relation with probability .25, and the overlapping set relationship with probability .75; although relations of equivalent sets and subset (*A*) - set (*B*) are also logically correct, they are never used. The probabilities are suggested by Erickson as approximations, not final values.

The combination process, like the encoding process, is incomplete. Subjects choose at random one of the possible ways of combining two representations (one from each premise). For example, suppose that _B_ and _C_ are overlapping sets, and _A_ is a subset of _B_. Then _A_ and _C_ might be disjoint; or they might be overlapping; or _A_ might be a subset of _C_. The subject chooses one of these relations at random in the combination process.

Complete-combination model. This model, also proposed by Erickson (1974), makes the same assumptions about encoding as does the random combination model. The difference between models lies in the combination stage. In this stage, subjects are assumed to combine the two representations (one from each premise) in a logically correct and complete manner, that is, in all possible valid ways. In the example given above, therefore, the subject would compute all three possible relations between _A_ and _C_.

Transitive-chain model. In this model, proposed by Guyote and Sternberg (Note 1), subjects are assumed to encode premises in a logically correct and complete manner. Errors can result from the combination stage, however. First, subjects are assumed to combine a maximum of four pairs of representations. This maximum is set by the limits of working memory. Consider, for example, the premises "Some _B_ are _C_. All _A_ are _B_." Complete encoding of the first premise results in four set relations: _B_ and _C_ equivalent, _B_ and _C_ overlapping, _B_ as a subset of _C_, _B_ as a superset of _C_. Complete encoding of the second premise results in two set relations: _A_ and _B_ equivalent, _A_ as a subset of _B_. There are eight possible ways to combine the first four set relations with the second two set relations. Subjects use a maximum of four, however. Second, identical representations--_A_ and _B_ equivalent--are combined before symmetrical representations--_A_ and _B_ overlapping, _A_ and _B_ disjoint--which are combined before asymmetrical representations--_A_ as a subset of _B_, _A_ as a superset of _B_. In the present context, a symmetrical relation is defined as one in which the positions of A and B could be reversed without changing the meaning of the representation. In an asymmetrical relation, the positions of A and B cannot be changed without changing the meaning of the relation.

Quantification of Models

The random-combination, complete-combination, and transitive-chain models were specified by their original authors in sufficient detail to permit quantification. Quantification of the ideal model is straightforward, since it always predicts the logically correct answer. The atmosphere and conversion models were not originally specified in sufficient detail to permit quantification, and so we have added what we believe to be minor assumptions that are consistent in spirit with the original models. Quantification procedures for each of the six models are described in the appendix, although reading of this appendix should be deferred until reading of the Method section is completed.

Inevitably, the numbers of free parameters will vary across the alternative models because of their radically different assumptions about the sources of error in syllogistic reasoning. At one extreme, the ideal, atmosphere, and conversion models as formulated in this article permitted predictions of responses without estimation of any parameters. At the other extreme, the random- and complete-combination models as formulated here required as many as nine parameters. The transitive chain model required four. The differences in numbers of parameters obviously must be taken into account in assessing comparative model fits. As it turned out, however, there was little relationship between the number of parameters in a model and its fit to the data. In view of this fact, and the fact that our primary goal was to determine which of the models best fit the data, regardless of the number of sources of error it implied, we were not particularly concerned about differences in numbers of parameters.

Method

Subjects

Subjects were 64 Yale College students participating for credit in an introductory psychology course. Thirty-seven of the subjects were men, and 27 were women. None of the subjects had ever had a course in formal logic.

Materials

Premises. Premises were presented in the context of either of two tasks, an encoding task or a combination task. In the encoding task, subjects would receive a premise such as "Some *A* are not *B*," and a conclusion such as "No *A* are *B*." The subject would have to indicate whether the conclusion was definitely, possibly, or never true of the single premise. In this particular case, the logically correct answer is "possibly true," since if at least some *A* are not *B*, it is possible that in fact none of the *A* are *B*. In the combination task, subjects would receive a pair of premises such as "Some *B* are *C*. All *B* are *A*," plus a conclusion, such as "Some *A* are *C*." The subject would have to indicate whether the conclusion was definitely, possibly, or never true of the premises considered together. In this particular case, the logically correct answer is "definitely true," since for the given premises, there will always be at least one *A* that is a *C*.

Premises in the encoding task used the letters *A* and *B*. Premises in the combination task used the letters *A*, *B*, and *C* to denote the subject, middle term, and predicate of each syllogism. All premises in both tasks were either affirmative or negative, and universal or particular, yielding four possible statements: All *A* are *B*, Some *A* are *B*, No *A* are *B*, Some *A* are not *B*. In the encoding task, these four statements served as the single premises presented to subjects. In the combination task, there were 15 different pairs of premises, 11 of which had at least one valid conclusion and 4 of which had no valid conclusion among the standard subset of four (described below). The pairs of premises are shown in Table 5.

Conclusions. Each premise or pair of premises was presented with only a single conclusion at a time. Over the course of the experiment, however, subjects received the premises together with either a full set of 10 conclusions or a standard subset consisting of 4 conclusions from the full set. The subset consisted of the conclusions commonly used in syllogistic reasoning tasks: All A are C, Some A are C, No A are C, Some A are not C. (In the encoding task, B was substituted for C.) The full set of conclusions consisted of the standard subset, plus six additional conclusions: All C are A, Some C are not A, All A are C and all C are A, Some but not all A are C and some but not all C are A, All A are C and some C are not A, All C are A and some A are not C. (In the encoding task, B was substituted for C.) The reason for including these particular additional conclusions will be discussed later.

Design

The basic design of the experiment was a two-by-two factorial arrangement, with task (encoding or combination) crossed with number of conclusions (standard subset or complete set). Each of the four conditions involved testing of 16 subjects, none of whom overlapped between conditions. The dependent variable was each subject's response of "definitely true," "possibly true," or "never true" to each conclusion for each premise or pair of premises.

Procedure

Subjects were tested in small groups. At the beginning of the testing session, subjects were told the nature of the task, and that we were concerned with logical validity of conclusions. Subjects were warned that "possibly true" was not intended as a weak form of, or hedge for, "definitely true." It was to be used only when the conclusion was neither definitely true nor never true of the premise(s). As is customary in experiments on syllogistic reasoning, subjects were informed that the logical meaning of some is "at least some and possibly all," which contrasts with the everyday meaning of some, "some but not all." After the instructions were completed, subjects tried one practice problem with each type of conclusion they would later receive in the test trials. Thus, subjects in the subset groups received four practice items, and subjects in the full-set groups received ten practice items.

When they were ready to begin the experimental task, subjects were told that problems must be solved in the (random) order in which they were presented, that they were not to refer back to previous problems in solving new problems, and that they should respond as accurately as possible. The subjects then began the solution of test problems, which were printed one to a page in computer-generated booklets. Subjects had as much time as they needed to complete the problems, which numbered 40 in the encoding task with the complete set of conclusions, 16 in the encoding task with the standard subset of conclusions, 150 in the combination task with the complete set of conclusions, and 60 in the combination task with the standard subset of conclusions.

Results

Basic Statistics

For the subset of conclusions common to both the group receiving the full set of conclusions (hereafter, the "full-set group") and the group receiving the subset of conclusions (hereafter, the "subset group"), the mean proportion of logically correct responses in the encoding task was .90 for the full-set group and .90 for the subset group. The mean proportion of logically correct responses in the combination task was .79 for the full-set group and .74 for the subset group. An analysis of variance revealed a highly significant effect of task, $F(1,60) = 21.41$, $p < .001$, but trivial effects of group, $F(1,60) < 1$ and task by group interaction, $F(1,60) < 1$.

These data indicate that the combination task is harder than the encoding task, which is to be expected since the combination task requires prior encoding of the premises of the problem. Performance on the encoding task is considerably better than would be expected from the predictions of most of the performance models, but not as good as the perfect performance predicted by the ideal and transitive-chain models. In order to evaluate performance on this task and the combination task more precisely, it is necessary to examine the results of the comparative model testing.

Comparative Model Testing

The alternative models were quantified according to the principles described earlier and using the procedures outlined in the appendix. The models were tested through model predictions derived in two ways. First, parameter estimates derived from the selection task for abstract syllogisms in Experiment 1 of Guyote and Sternberg (Note 1) were entered into the appropriate equations for the present evaluation task, and used as the basis for predicting responses Second, parameters were estimated from the present data sets and used as the basis for predicting responses. The first procedure enabled us to test the stability and generality of parameter estimates over subjects (with two samples drawn from the same population) and over tasks (response selection in Guyote & Sternberg [Note 1] versus response evaluation in the present experiment; combination in Guyote & Sternberg [Note 1] and the present experiment versus encoding in the present experiment only). Since the number of data points was also greater in the Guyote and Sternberg experiment than in the present experiment (180 versus a range of 16 to 150--see Tables 1 and 2), the parameter estimates from the selection task could be expected to be somewhat more constrained than those from the present evaluation task. The second procedure enabled us to test how well the models fit the data when parameters were estimated specifically for the tasks and subjects at hand.

Encoding. Table 1 shows comparative model fits for the encoding task.

Insert Table 1 about here

Model fits are presented for three data sets. The first data set is for the standard subset of conclusions (All A are C; No A are C; Some A are C; Some A are not C) as responded to by the subset group. The second data set is for the same subset of conclusions as responded to by the full-set group. The third data set is for the full set of conclusions as responded to by the full-set group. Model fits are presented in terms of the proportion of variance in the data accounted for by each model (R^2) and the root-mean-square deviation of the observed from the predicted values (RMSD). Each data point was a mean over subjects of the proposed response (1= definitely true, 2=possibly true, 3=never true) of the subjects to each conclusion for each of the four premises.

The reliability of each data set was computed by arbitrarily dividing the subjects into two groups of equal size, computing the correlation across item types between quantified responses for the two sets of subjects, and correcting this correlation by the Spearman-Brown formula to take into account the fact that only half the subjects were used for computing the values of each variate.

With the Guyote-Sternberg parameter estimates, the ideal-subject and transitive-chain models do an excellent job of predicting subjects' responses for all three data sets. The complete- and random-combination models and the atmosphere model clearly do not predict the data as well, despite the large number of parameters estimated for the first two models. The conversion model makes predictions indistinguishable from those of the ideal-subject and transitive-chain models for the subset of conclusions, but its predictions are distinguished for the full set of conclusions; and for this set, the conversion model is inferior to the ideal-subject and transitive-chain models. These data are therefore consistent with the notion that subjects approximate a strategy of complete and correct encoding of the syllogistic premises.

The parameters as estimated directly from the present data are less useful in distinguishing among models. Possible reasons for this lessened distinguishability are discussed in the next section. Whatever the reasons, it appears that the larger numbers of parameters in the complete- and random-combination models buy virtually nothing in the present analysis in terms of improved prediction of the complete- and random-combination models over the ideal-subject and transitive-chain models.

It is important to note that not even these two preferred (and indistinguishable) models for the encoding task perform at a level comparable to the reliability of the data. Thus, the ideal-subject/transitive-chain strategy of complet and correct encoding can be viewed only as an approximation to the strategy or strategies subjects actually use. A true model of performance in the encoding task would have to account for all of the systematic variance in the data, something none of the models tested is able to do.

Is nearly flawless encoding characteristic of individual as well as group performance, or might it be some averaging artifact that does not in fact veridically represent the performance of individual subjects? The answer to this question was sought through the modeling of individual data. Table 2 shows individual model fits for members of the subset group. Fits for each model are presented both in terms of R^2 and RMSD. Because the models predict probabilities of responses, parameter estimation for individual subjects would be possible only with very large numbers of replications. Since there were no replications for individual subjects in this experiment, the group parameter estimates (from Experiment 1 of Guyote & Sternberg, Note 1) were used in the model fitting. The absence of replications also made it impossible to compute reliability of the individual data in the standard fashion. The reliability index (or square root of the reliability coefficient) was therefore estimated as the correlation between an individual's pattern of data and the group pattern of data (with that individual excluded). This estimate of the reliability index is conservative, since it assumes that the group pattern represents the true pattern (over infinite replications) for each individual.

Insert Table 2 about here

The first thing to note in the table is that despite the absence of replicated observations for individual subjects, the data for these subjects were highly reliable. The estimated reliability index was at least .85 in 15 of 16 cases, and at least .90 in 11 of 16 cases. Subject 10, with an index of .58, was an anomaly.

The second thing to note is that a model of flawless encoding best accounted for the data of almost all of the subjects (13 of 16 using R^2 as the criterion, or 14 of 16 using RMSD as the criterion). This model accounted for all of the variance in the data of 6 of 16 subjects, and the mean R^2 across subjects was .84, a most respectable level of fit for individual data. This mean is reduced considerably by the anomalous Subject 10, whose data were of low reliability and for whom none of the models were satisfactory. (Note that for the subset of conclusions, the predictions of the conversion model cannot be distinguished from those of the ideal and transitive-chain models. The data for the full set of conclusions does distinguish models, and the individual data for the full set of conclusions, like the group data, argue strongly against the conversion model.)

In general, the patterns of RMSD closely follow those of R^2. This congruence of outcome is to be expected where subjects use the scale of response choices in much the same way, since the major difference between R^2 and RMSD is that R^2 allows an additive constant in evaluating fit, whereas RMSD does not. Individual differences in the use of the response scale (1=definitely true, 2=possibly true, 3=never true) appear to have been quite small. Mean scale values averaged across items ranged from 1.94 to 2.31 for individual subjects, with a grand mean of 2.04 and a standard deviation of .08. The ideal subject would have shown a mean of 2.00. Although subjects differed little among themselves, and 9 of 16 subjects had means of exactly 2.00, the mean across subjects (2.04) did differ significantly from the ideal mean (2.00), $\underline{z} = 2.00$, $\underline{p} < .05$. Subjects were on the average slightly more conservative than the ideal subject would be, as would be expected if they sometimes failed to apprecia[illegible] all possible encodings of a premise. Standard deviations of responses for individual subjects ranged from .73 to .93, with a mean across subjects of .87 and a standard deviation of .05. This mean standard deviation was significantly lower than that which would be shown by the ideal subject (.89), $z = -2.15$, $p < .05$, indicating that subjects were less variable in their responses than an ideal subject would have been.

Combination. Table 3 shows comparative model fits for the combination task. Reliability of the data was computed in the same way as for the encoding task. Model testing based upon the Guyote-Sternberg parameter estimates

---

Insert Table 3 about here

---

once again indicates the superiority of the transitive-chain model over the alternative models. This model has the highest value of R^2 and the lowest value of RMSD for each of the three data sets.[2] The absolute levels of fit are also quite respectable, especially for the subset of conclusions. As in the encoding task, however, the values of R^2 are lower than the reliability, indicating the presence of systematic variance unaccounted for by any of the models. None of the models is true, therefore, in accounting for subjects' strategies in encoding and combining premise information.

As was the case in the analyses for the encoding task, the models are less distinguishable when parameters are estimated for the present data. Why do parameter estimates from the evaluation task consistently distinguish models less well than the estimates from the selection task? There seem to be three possible reasons, any or all of which may contribute to lessened distinguishability in the evaluation task. First, the number of data points from which parameters were estimated was larger in the Guyote-Sternberg experiment than in the present experiment, so that there was less opportunity for capitalization upon chance in the earlier experiment. In models with relatively large numbers of parameters (such as the complete- and random-combination models), capitalization upon chance can be a serious problem. Second, the response-evaluation task used in the present experiment eliminates the comparison stage of syllogistic reasoning, which is a major source of variation in the differential predictions of the alternative models. Third, there seems to be at least some difficulty in the combination task in distinguishing between relations that have not been encoded and those that have been encoded but not combined.[3] The encoding task, therefore, is useful in showing the very high level of performance subjects reach in encoding when encoding is isolated from combination.

Table 4 shows individual model fits for members of the subset group. Procedures for computing the individual model fits and reliability indices were the same as in the analysis of the individual encoding data.

Insert Table 4 about here

The data are again highly reliable: The mean correlation between individual and group data was as high as .84. The transitive-chain model was again the preferred model, best accounting for the data of 7 of 16 subjects with R^2 as the criterion, or 8 of 16 subjects with RMSD as the criterion. The absolute levels of fit, with mean $R^2 = .72$ and mean RMSD = .44, were also quite respectable for individual-subject data. In all but two cases where a model other than the transitive-chain model was preferred, the difference in R^2 and RMSD between the transitive-chain model and the preferred model was small (between .01 and .03 units of RMSD). In two cases, however (Subjects 9 and 11), the individual data were clearly better fit by the ideal-subject model than by any other model. It thus appears that a small minority of subjects are able not only to encode premises almost perfectly, but to combine them almost perfectly as well. On the whole, however, the individual data argue in favor of the general superiority of the mixed model over its competitors.

Variation in use of the response scale was relatively small, although not as small as in the encoding task. Mean responses for individual subjects ranged from 1.85 to 2.23, with a mean of 2.02 and a standard deviation of .085. Standard deviations (representing variability in responses) ranged from .70 to .92, with a mean of .80 and a standard deviation of .073.

Parameter Estimates

Comparison across subjects and tasks. One basis for assessing the adequacy of a model is the stability of the model's parameter estimates across subjects and task variants. Stable parameter estimates indicate that the model is generalizable across subjects and tasks, and that capitalization upon chance is not a major factor in determining fit of the model for any one task or group of subjects. Unstable parameter estimates, on the other hand, call into question the generalizability of the model and the interpretability of any one set of parameter estimates.

Table 5 shows parameter estimates for the three models for which parame-

---

Insert Table 5 about here

---

ters were estimated. Since encoding is assumed to be complete and correct in the transitive-chain model, there are no parameters associated with the encoding stage of this model, and parameters were estimated for the combination task only. Our first concern was to assess the stability of the parameter estimates across subjects and tasks. This assessment was made by computing the root-mean-square deviation of the estimates across subjects and tasks. These RMSDs are thus across pairs of columns for a given model in Table 5. For the transitive-chain model, RMSD is .15 between the Guyote-Sternberg and Sternberg-Turner data. RMSDs were higher for the alternative models: For the complete-combination model, the corresponding value of RMSD is .22, and for the random-combination model, it is .36. Thus, the parameter estimates of the transitive-chain model do appear to be more stable and generalizable across subjects and tasks, perhaps because of the reduction in chance fluctuations often associated with estimation of a smaller number of parameters. The values of RMSD were also computed between the encoding and combination tasks in the present experiment. The values of RMSD were .15 for the complete-combination model and .20 for the random-combination model.

Interpretation of parameter estimates. The value of p_1 indicates that subjects combine only one pair of set relations about half the time, and combine more than one $(p_2+p_3+p_4)$ the other half. This restriction in the number of combinations performed is assumed to be due to the limitations of working memory.

The parameters of the complete- and random-combination models signify probabilities of encoding different premises by means of various set relations (see Table 5). For example, the probability of encoding "All A are B" as A equivalent to B is estimated to be .80 in the present data for the complete-combination model. Erickson's model as originally formulated contained only what are here called p_1, p_2, p_3, and p_4. One of these parameters, p_3, is estimated as 0 in all but one data set, where it is estimated as .01. The augmentation of the models proposed here considerably improves the predictive power of the models, however, in that most of parameters p_4-p_9 have values that depart substantially from either 0 or 1. Apparently, "Some A are B" and "Some A are not B" do have a variety of interpretations assigned to them, if one accepts either of these models as valid (which we do not).

Modified Truth Table Analysis

The global quantitative analyses of model fits and parameter estimates do not enable one to assess qualitative features of the data, such as the nature of the representations subjects use. We have examined these qualitative features of the data through a modified truth-table analysis similar in some respects to those employed by Staudenmayer (1975) and by Taplin, Staudenmayer, and Taddonio (1974) for conditional premises. However, our methodology in conducting this analysis was quite different from anything that has been attempted before.

Consider as an illustration the premises "All B are C. All A are B." We wish to know how subjects represent the set relation(s) that characterize(s) (a) each of the premises and (b) the combined premises. We will deal in this illustration only with item (b), since the principles are the same for item (a). There are five possible set relations relating A to C: A equivalent to C, A superset of C, A subset of C, A overlapping with C, A disjoint with C. A subject's combined representation may contain any of the $2^5-1 = 31$ possible nonnull subsets of these five set relations: If a subject has a logically consistent (but not necessarily correct) representation, it will be one of these 31 possible ones. Obviously, some combined representations are much more plausible than others. In the present example, three plausible combined representations might be "A equivalent to C," "A subset of C," and "A equivalent to C or A subset of C." (The last representation is logically complete and correct.) Is there some way of finding out which of these (or other) representations subjects actually used? Our proposed way is through the modified truth-table analysis.

Consider an example of such an analysis presented in Table 6. Each of

---

Insert Table 6 about here

---

the 10 conclusions for the full-set group is shown at the left, and three columns are shown at the right, one for each of the three plausible representations. It turns out that each of these three representations predicts a different pattern of "definitely true," "possibly true," and "never true" responses to the 10 conclusions given to the full-set group, and in fact, each of the 31 possible representations predicts a different pattern. The 10 conclusions were chosen so as to yield a unique pattern of responses that would permit us to distinguish the best (most frequently used) of the 31 possible representations.

Suppose that a subject believes the correct combined representation to be that of "A equivalent to C." (He or she does not realize that "A subset of C" is also possible.) Then the conclusion "All A are C" is definitely true of this representation. This conclusion is also definitely true of each of the other three plausible representations of the combined set relation, so that the conclusion is not helpful in distinguishing which representation the subject actually uses. In fact, all four of the standard syllogistic conclusions (All A are C, No A are C, Some A are C, Some A are not C) given to the subset group yield the same pattern of predictions, so that one could not distinguish among the three plausible representations on the basis of responses to these conclusions: "All A are C" is definitely true of all of them; "No A are C" is never true of any of them; "Some A are C" is definitely true of all of them; and "Some A are not C" is never true of any of them. Consider, however, the conclusion, "All C are A." This conclusion leads to a different response for each of the plausible representations. It is definitely true of "A equivalent to C." It is never true, however, of "A subset of C," since in this conclusion, some but not all of the C are A. It is possibly true of the third representation, in that it is true of one-half of the representation (A equivalent to C) but not of the other half _ subset of C).

We computed the root-mean-square deviation between the pattern of responses to the 10 conclusions predicted by each of the 31 logically consistent representations (again using 1=definitely true, 2=possibly true, 3=never true) and both (a) the observed pattern of responses, and (b) the pattern of responses predicted by each of the models of syllogistic reasoning. Parameter estimates from the present experiment were used in making these predictions. It was possible through these computations to infer for the full-set group both the representations they actually used for each syllogism and the representation they were alleged to use under the assumptions of each model. It seems likely that there would have been individual differences in the representations used by different subjects, and even in the representations used by single subjects at different times. We examined only mean data, however, because of its greater stability and because the extremely high internal-consistency reliability of the group data for split halves of subjects suggested that individual differences were not a major influence upon the results.

Table 7 shows the representation with the lowest RMSD for each of the encoding and combination problems. The models differed in the accuracies

Insert Table 7 about here

with which they predicted these representations. The average rank order of the best-fitting representation (as determined by the actual data) among the 31 possible representations was computed for each model for both the encoding and combination tasks. A rank of 1, for example, would indicate that the best representation according to the model is also the one the data indicate subjects use. A rank of 2 would indicate that the actual representation was ranked only second best by the model. The average rank orders for the respective tasks were 1.00 and 1.40 for the ideal-subject model, 1.00 and 1.40 for the transitive-chain model, 1.25 and 2.47 for the complete-combination model, 1.25 and 2.60 for the random-combination model, 1.50 and 2.20 for the conversion model, and 5.00 and 6.47 for the atmosphere model. These results indicate that the ideal-subject and transitive-chain models were better than the alternative models in predicting the best-fitting representation, although these two models were not distinguishable, possibly because subjects receiving the full set of conclusions had more of an opportunity to reflect upon possible representations for the combined premises, and thus may have more closely resembled ideal subjects.

Discussion

The present experiment investigated subjects' performance in two tasks, one requiring only encoding of the premises of syllogisms, the other requiring both encoding and combination of premises. A major finding was that subjects made only about 10% errors in the encoding task. Since this figure includes constant sources of error as well as error due specifically to encoding, the result suggests that errors in encoding are not a major source of difficulty in syllogistic reasoning. Modeling of the encoding data in the present experiment also supports the notion that encoding is not a major source of response-choice variation in the syllogisms task. The ideal-subject and transitive-chain models, which predict error-free encoding, performed better than the alternative models that were considered. The differentiation among models showed up clearly only when the Guyote-Sternberg parameter estimates were used, presumably in part because of the greater constraints in these earlier data. To summarize, then, the present data suggest that encoding is nearly error-free, and that most errors occur in combination (and in the response-selection task, in comparison as well).

This conclusion is opposed to that of certain other authors, most notably Ceraso and Provitera (1971). Like us, Ceraso and Provitera claimed to test the encoding hypothesis directly. But whereas we sought to test the encoding hypothesis by decomposing the syllogistic evaluation task and by manipulating the conclusions given to the full-set group, Ceraso and Provitera sought to test the hypothesis by manipulating the nature of the premises given to a "modified" syllogisms group. The authors proposed that most errors in syllogisms tasks are attributable to encoding failure rather than to reasoning errors. They had subjects solve either traditional syllogisms (using a variant of the response-selection task) or modified syllogisms. The latter syllogisms were "modified" in the sense that subjects were taught to represent each premise by just one set relation. Previous investigations by Ceraso (Note 2) had suggested that many people in fact represent the premises of traditional syllogisms only by single set relations.

Ceraso and Provitera interpreted their results as supporting the encoding hypothesis. This interpretation bears further examination. It appears that the performance of subjects in the modified group serves only as a poor model for the performance of subjects in the traditional group: The squared correlation between response-choice probabilities of subjects in the two groups was only .69. One is not safe in concluding, therefore, that subjects "responded to the traditional syllogisms as if they were the modified syllogisms, which accounted for their errors" (Ceraso & Provitera, 1971, p. 400).

How well would the transitive-chain model fit Ceraso and Provitera's data? Of the 13 syllogisms used by Ceraso and Provitera, 11 were used in Experiment 1 of Guyote and Sternberg (Note 1). We used the parameter estimates from Guyote and Sternberg (Note 1) for the 11 common syllogisms to predict Ceraso and Provitera's data for the traditional-syllogisms group. Since Ceraso and Provitera did not have as a conclusion "Some A are not C," we combined this conclusion with "Some A are C," which could be expected slightly to reduce the fit of the transitive-chain model. Recall also that Ceraso and Provitera did not teach subjects to use the logical meaning of some, whereas Guyote and Sternberg did. This difference could also be expected to reduce the fit. Nevertheless, the squared correlation between the predictions of the transitive-chain model for Guyote and Sternberg's data and Ceraso and Provitera's traditional-group data was .91. The squared correlation between the modified-group and traditional-group data for these same 11 syllogisms was .69 (the same as for the full set of 13 syllogisms). Thus, the predictions of the transitive-chain model for different subjects in a different experiment administered with different instructions were considerably better than predictions based upon the modified-group data.

Ceraso and Provitera, recognizing that the traditional-group data were not well predicted by the modified-group data, suggested two other sources of error in syllogistic reasoning. The first was atmosphere of the premises. The second was incomplete combination of set relations. Both of these sources of error are predicted by the transitive-chain model of syllogistic reasoning, according to which working memory limitations result in incomplete information processing during the combination stage, and atmosphere of the premises serves as one heuristic for choosing a preferred conclusion during the comparison stage (see Guyote & Sternberg, Note 1).

According to Ceraso and Provitera, a subset of five of the syllogisms they used provided a particularly important test of their model, in that neither of the two additional sources of error noted above should have operated differentially for the traditional- and modified-syllogisms groups. For these five syllogisms, the squared correlation between response-choice patterns in the two groups was .92. Three of these syllogisms were also used by Guyote and Sternberg (Note 1). The squared correlation between the predictions of the transitive-chain model for the Guyote-Sternberg data and Ceraso and Provitera's response-choice data was .96. The comparable squared correlation for these three syllogisms, using the modified-group data to predict the traditional group data, was also .96.

We conclude from these analyses that the transitive-chain model provides as good an account as Ceraso and Provitera's encoding hypothesis for the small subset of the data considered immediately above, and provides a much better account of the data as a whole than does the encoding hypothesis taken alone. Taken as a whole, therefore, the results of the present experiment as well as Ceraso and Provitera's are consistent with the notion that encoding errors are not a major source of response-choice variation in syllogistic reasoning. If subjects are not instructed to use the logical meaning of some, it seems likely that response-choice variation due to the encoding stage may increase. But this increased variation, to whatever extent it may exist in Ceraso and Provitera's and other experiments, seems likely to be due not to an (interesting) inability to translate verbal descriptions into set relations, but to a (less interesting) ignorance on the part of subjects of what a particular verbal description is supposed to mean in sentential logic.

Modeling of the combination data also supported the transitive-chain model, at least for the standard subset of conclusions. The data from the subset of conclusions seem to rule out the ideal-subject, conversion, and atmosphere models relative to the transitive-chain model. The complete- and random-combination models are ruled out only if the Guyote-Sternberg parameter estimates are used. The transitive-chain model, however, showed more stable parameter estimates across subjects and tasks than did either the complete- or random-combination models.

Using a modified truth-table analysis, we were able to infer the representations used by full-set subjects in encoding and combining premises. It was noted, however, that the additional conclusions given to full-set subjects in order to enable us to infer their representations may have resulted in their becoming aware of additional possible set relations.

Although the results of this experiment and the previous ones (Guyote &

Sternberg, Note 1) generally support the transitive-chain model, there is systematic response-choice variation left unaccounted for in all of the experiments conducted so far. Neither the transitive-chain model, nor any other model considered, for that matter, is true. The transitive-chain model is no doubt a simplification of the model subjects actually use. First, it does not fully account for performance during encoding. The evidence suggests that although encoding is not nearly so flawed as some investigators have believed, neither is it flawless, as assumed by the transitive-chain model. Second, the model does not fully account for the effects of syllogistic figure (the order in which the terms appear in each of the premises) upon response choice. Previous evidence (for example, Dickstein, 1978) suggests that figure has fairly complex effects upon subjects' choices of responses. A more sophisticated model, therefore, might take into account all of these complex effects. Attempts at the formulation of such a model are presently being made.

Appendix

Quantification of Models of Syllogistic Reasoning

In this appendix, the quantification of the models of syllogistic reasoning will be described in some detail. Each description will be based upon an example from the combination task, All _C_ are _B_. Some _A_ are not _B_.

Ideal Model

Encoding. According to this model, each premise is encoded completely and correctly. Thus, the first premise of the example syllogism is represented as

1a. _C_ equivalent to _B_.

1b. _C_ subset of _B_.

and the second premise is represented as

2a. _A_ overlapping with _B_.

2b. _A_ disjoint with _B_.

2c. _A_ superset of _B_.

Combination. The model also assumes the the premises are combined completely and correctly. Each of the two set relations from the first premise can be combined with each of the three set relations from the second premise, yielding six combinations.

1a & 2a. _A_ overlapping with _C_.

1a & 2b. _A_ disjoint with _C_.

1a & 2c. _A_ superset of _C_.

1b & 2a. _A_ overlapping with _C_ or _A_ disjoint with _C_ or _A_ superset of _C_.

1b & 2b. _A_ disjoint with _C_.

1b & 2c. _A_ superset of _C_.

The representations resulting from the various combinations are obviously non-unique. In fact, there are exactly three distinct set relations between _A_ and _C_ _A_ overlapping with _C_, _A_ disjoint with _C_, _A_ superset of _C_.

Response. Responses to the various conclusions are chosen on the basis of this combined representation. Consider as examples the four conclusions from the standard subset. "All A are C" fails to describe any of the three composite set relations, and so is never true. "No A are C" correctly describes one of the three composite set relations (A disjoint with C), and so is possibly true. "Some A are C" correctly describes two of the three relations (A overlapping with C, A superset of C), and so is also possibly true. "Some A are not C" correctly describes all three of the relations, and so is definitely true. In quantifying predicted and observed values, a response of "definitely true" was assigned a numerical value of 1; a response of "possibly true" was assigned a numerical value of 2; a response of "never true" was assigned a numerical value of 3.[4]

Transitive-Chain Model

Encoding. The transitive chain model, like the ideal model, assumes complete and correct encoding, so that the discussion of encoding for the ideal model is relevant here as well.

Combination. Encodings 1a, 2a, and 2b are symmetrical, and hence preferred. As a result, there are two preferred pairs of set relations--1a & 2a, 1a & 2b. These will always be combined first. Let

p_1 = P(performing one combination)

p_2 = P(performing two combinations)

p_3 = P(performing three combinations)

p_4 = P(performing four combinations)

If just one combination is performed, it will be with one of the two preferred pairs of set relations:

1a & 2a. A overlapping with C.

or

1a & 2b. A disjoint with C.

Since each of the two possible combinations is equally likely,

P(A overlapping with C) = $1/2(p_1)$

P(A disjoint with C) = $1/2(p_1)$

If two combinations are performed, they will be the two preferred ones, so that

P(A disjoint with C, A overlapping with C) = p_2

Suppose that three combinations are performed. They will be some subset of the six possible combinations of 1a and 1b with 2a, 2b, and 2c. Six combinations taken three at a time results in 20 possible subsets. But the two preferred pairs must be included, so that only one of the four combinations is left to vary. Four things taken one at a time results in four possible subsets. Listing of the elements in these subsets reveals that

P(A disjoint with C, A overlapping with C, A superset of C) = $3/4(p_3)$

P(A disjoint with C, A overlapping with C) = $1/4(p_3)$

Suppose, finally, that four combinations are performed. The same logic applies as was used above in determining possible combinations. There end up being six possible subsets. Each of the subsets yields the same three composite representations, so that

P(A disjoint with C, A overlapping with C, A superset of C) = p_4

To summarize, subjects might use any of four possible representations, depending upon how many and which combinations they performed. The four possible representations and their associated probabilities are

P(A disjoint with C) = $1/2(p_1)$

P(A overlapping with C) = $1/2(p_1)$

P(A disjoint with C, A overlapping with C) = $p_2 + 1/4(p_3)$

P(A disjoint with C, A overlapping with C, A superset of C) = $3/4(p_3) + p_4$

<u>Response</u>. The predicted response for each conclusion can now be computed. Let us consider the prediction for the conclusion, "Some <u>A</u> are <u>C</u>." If the subject uses the first representation (<u>A</u> disjoint with <u>C</u>), the conclusion is never true. If the subject uses the second representation (<u>A</u> overlapping with <u>C</u>), the conclusion is definitely true. Finally, if the subject uses the third or fourth representation (<u>A</u> disjoint with <u>C</u>, <u>A</u> overlapping with <u>C</u> or <u>A</u> disjoint with <u>C</u>, <u>A</u> overlapping with <u>C</u>, <u>A</u> superset of <u>C</u>), the conclusion is possibly true, since at least one, but not all of the set relations in each representation can be appropriately described by this conclusion. We are now able to state that

$$P(\text{"definitely true"}) = 1/2(p_1)$$

$$P(\text{"possibly true"}) = p_2 + p_3 + p_4$$

$$P(\text{"never true"}) = 1/2(p_1)$$

It is now possible to compute the predicted value for the conclusion, keeping in mind that "definitely true" is assigned a numerical value of 1, "possibly true" a numerical value of 2, and "never true" a numerical value of 3. The predicted value is simply the expected value of the possible responses:

$$\begin{aligned}\text{Predicted value} &= 1[1/2(p_1)] + 2[p_2 + p_3 + p_4] + 3[1/2(p_1)]\\ &= 2(p_1 + p_2 + p_3 + p_4) = 2\end{aligned}$$

In this particular example, it is not necessary even to estimate the individual values of p_1, p_2, p_3, and p_4, since they must sum to 1. In other examples, however, the values of the probabilities must be estimated in order to predict the response to a given conclusion.

<u>Random-Combination Model</u>

<u>Encoding</u>. According to Erickson (1974), a given set relation may have a predefined probability of zero or one of being used to encode a given premise, or it may have an estimated probability of being used. Consider the premise "All <u>A</u> are <u>B</u>." Two parameters, which we shall label p_1 and p_2, are associated

with encoding of this premise:

p_1 = P(encoding "All A are B" as A equivalent to B)

p_2 = P(encoding "All A are B" as A subset of B)

All other encodings are used with probability zero. The premise "No A are B" is encoded as A disjoint with B with probability one. In Erickson's formulation of the random-combination model, two parameters, which we shall call p_3 and p_4, are associated with the encoding of the premise "Some A are B":

p_3 = P(encoding "Some A are B" as A superset of B)

p_4 = P(encoding "Some A are B" as A overlapping with B)

All other encodings are used with probability zero. Erickson apparently did not inform his subjects of the logical meaning of "some," as is commonly done in experiments on syllogistic reasoning. We did inform subjects of this meaning, however, rendering it implausible that subjects would never use the other two set relations that properly describe the premise "Some A are B." We therefore estimated two additional parameters for our data:

p_5 = P(encoding "Some A are B" as A equivalent to B)

p_6 = P(encoding "Some A are B" as A subset of B)

The premise "Some A are not B" is proposed by Erickson to be encoded as A overlapping with B with probability one. In order to maximize the fit of the model to the data, three additional parameters were also estimated, corresponding to probabilities of encoding "Some A are not B" in various ways:

p_7 = P(encoding "Some A are not B" as A superset of B)

p_8 = P(encoding "Some A are not B" as A overlapping with B)

p_9 = P(encoding "Some A are not B" as A disjoint with B)

Note that the model with parameters as estimated is an augmented version of Erickson's original model. Its performance will always be equal to or better than that of the original model, since fixed constants in the original model are now estimated as free parameters. The augmented version of the model retains Erickson's basic information-processing framework while permitting more flexibility in the encoding process.

Turning now to our example syllogism, we see that the first premise of the example problem can be encoded in either of two ways (1a or 1b), and the second premise can be encoded in any of three ways (2a, 2b, 2c). The probability that the first premise of the example will be encoded as C equivalent to B is p_1; the probability that it will be encoded as C subset of B is p_2. The probability that the second premise of the example will be encoded as A superset of B is p_7; the probability that it will be encoded as A overlapping with B is p_8; and the probability that it will be encoded as A disjoint with B is p_9. Note that in Erickson's model, subjects always encode each premise with only one set relation.

Combination. Since the first premise can be represented in two ways and the second premise in three ways, there are six possible ways of combining information from the two premises, which are the same as the six shown for the ideal model (see Combination section). The probabilities of the various combinations are 1a with 2a, p_1p_7; 1a with 2b, p_1p_8; 1a with 2c, p_1p_9; 1b with 2a, p_2p_7; 1b with 2b, p_2p_8; 1b with 2c, p_2p_9. In all of these possible combinations except 1b with 2a, there is only one possible resulting set relation, so that the probability of using each set relation is equal to the probabilities of the various combinations. If 1b is combined with 2a, however, there are three possible resulting set relations. In the random combination model, only one of the three combined set relations is constructed, however, and this one is chosen at random. Thus, the probability of any one of the three set relations being constructed is $1/3(p_2p_7)$.

We can now compute the probability that any given set relation will result from the combination process:

$P(\underline{A} \text{ overlapping with } \underline{C}) = p_1p_7 + 1/3(p_2p_7)$

$P(\underline{A} \text{ disjoint with } \underline{C}) = p_1p_8 + p_2p_8 + 1/3(p_2p_7)$

$P(\underline{A} \text{ superset of } \underline{C}) = p_1p_9 + 1/3(p_2p_7) + p_2p_9$

Response. The predicted value of the subject's typical response for each conclusion is computed in the same way as for the transitive-chain model. The predicted response is an expected value computed by summing the products of the numerical value of each possible response (definitely true, possibly true, never true) times the probability that each response is given. Consider the conclusion "Some A are C." The response "definitely true" will be given if either the set relation A overlapping with C or the set relation A superset of C is used. The response "possibly true" will never be given in this or any other example, since only one set relation is ever used in the combined representation. The response "never true" will be given if the set relation A disjoint with C is used. The associated probabilities are

$P(\text{"definitely true"}) = p_1p_7 + 2/3(p_2p_7) + p_1p_9 + p_2p_9$

$P(\text{"possibly true"}) = 0$

$P(\text{"never true"}) = p_1p_8 + p_2p_8 + 1/3(p_2p_7)$

The predicted value of the response to the conclusion "Some A are C" is thus

$$\text{Predicted value} = 1[p_1p_7 + 2/3(p_2p_7) + p_1p_9 + p_2p_9] + 2[0] + 3[p_1p_8 + p_2p_8 + 1/3(p_2p_7)]$$

Complete Combination Model

Encoding. This model makes the same assumptions about encoding as does the random combination model.

Combination. This model differs from the random combination model in one fundamental respect. If a combined representation comprises more than one set relation, all set relations are constructed rather than one being chosen at random. In the example problem, therefore,

$$P(A \text{ overlapping with } C) = p_1p_7 + p_2p_7$$

$$P(A \text{ disjoint with } C) = p_1p_8 + p_2p_8 + p_2p_7$$

$$P(A \text{ superset of } C) = p_1p_9 + p_2p_7 + p_2p_9$$

Response. The predicted response is computed in the same way as above. For the conclusion "Some A are C," the probabilities associated with each possible response are

$$P(\text{"definitely true"}) = p_1p_7 + p_1p_9 + p_2p_9$$

$$P(\text{"possibly true"}) = p_2p_7$$

$$P(\text{"never true"}) = p_1p_8 + p_2p_8$$

The predicted value of the response to the conclusion "Some A are C" is

$$\text{Predicted value} + 1(p_1p_7 + p_1p_9 + p_2p_9) + 2(p_2p_7) + 3(p_1p_8 + p_2p_8)$$

Conversion Model

Encoding. It is assumed that eachpremise is represented as the conjunction of the premise in its stated form and the premise in its converted form. Thus, the premise "All C are B" in the example is interpreted as meaning that "All C are B and all B are C," and is represented as C equivalent to B. The premise "Some A are not B" is interpreted as meaning that "Some A are not B and some B are not A," and is represented as A overlapping with B, A disjoint with B. Note that under this interpretation of the conversion model, representations of both premises are different from those obtained under the ideal model.

Combination. Combination proceeds as in the ideal model, except that combination is performed upon the representations as encoded by the rule stated above. Combining the one set relation from encoding of the first premise with the two set relations from encoding of the second premise yields the combined representation A overlapping with C, A disjoint with C.

Response. It is assumed that combined representations are compared directly to the conclusions, which are interpreted in unconverted form. Consider the conclusion "Some A are C." The predicted value for the response is 2 ("possibly true"), since one but not the other combined set relation is properly described by this conclusion.

Atmosphere Model

Encoding. According to this model, subjects are assumed to encode two features of each premise, the quantification (universal or particular) and the polarity (affirmative or negative). In the example syllogism, the first premise is universal affirmative and the second premise is particular negative.

Combination. Combination proceeds according to two rules. First, if either or both premises is particular, then the atmosphere of the syllogism is particular; otherwise, the atmosphere is universal. Second, if either or both premises is negative, then the atmosphere of the syllogism is negative; otherwise, the atmosphere is affirmative. In the example syllogism, the second premise is both particular and negative; hence, the atmosphere is particular negative.

Response. The subject evaluates the atmosphere of each conclusion. If the atmosphere of the combined premises matches the atmosphere of the conclusion in both quantification and polarity, the subject responds "definitely true." If the atmosphere of the combined premises matches the atmosphere of the conclusion in either quantification or polarity, but not both, the subject responds "possibly true." If the atmosphere of the combined premises matches the atmosphere of the conclusion in neither quantification nor polarity, the subject responds "never true." Consider, for example, the conclusion "Some A are C." This conclusion is particular affirmative. Since its atmosphere matches the atmosphere of the combined premises in quantification but not polarity, the subject is predicted to responds "possibly true," and the predicted value for this conclusion is therefore 2.

Reference Notes

1. Guyote, M. J., & Sternberg, R. J. A transitive-chain theory of syllogistic reasoning. (ONR Technical Report #5). New Haven, Conn.: Department of Psychology, Yale University, April, 1978. Also manuscript submitted for publication, 1978, and available from authors.

2. Ceraso, J. The effect of differing interpretations of sentences on the ability to use these sentences in the solution of syllogisms. Unpublished predoctoral thesis, New School for Social Research, 1957.

References

Ceraso, J., & Provitera, A. Sources of error in syllogistic reasoning. Cognitive Psychology, 1971, 2, 400-410.

Chapman, L. J., & Chapman, J. P. Atmosphere effect re-examined. Journal of Experimental Psychology, 1959, 58, 220-226.

Dickstein, L. S. The effect of figure on syllogistic reasoning. Memory and Cognition, 1978, 6, 76-83.

Erickson, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (Ed.), Loyola symposium on cognition (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Johnson-Laird, P. H., & Steedman, M. The psychology of syllogisms. Cognitive Psychology, 1978, 10, 64-99.

Revlis, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. Falmagne (Ed.), Reasoning: Representation and process. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Staudenmayer, H. Understanding conditional reasoning with meaningful propositions. In R. Falmagne (Ed.), Reasoning: Representation and process. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Sternberg, R. J. Isolating the components of intelligence. Intelligence, 1978, 2, in press.

Taplin, J. E., Staudenmayer, H., & Taddonio, J. L. Developmental changes in conditional reasoning: Linguistic or logical? Journal of Experimental Child Psychology, 1974, 17, 360-373.

Woodworth, R. S., & Sells, S. B. An atmosphere effect in formal syllogistic reasoning. Journal of Experimental Psychology, 1935, 18, 451-460.

Footnotes

Portions of this article were presented at the Mathematical Psychology Meetings in Los Angeles, August, 1977. We are grateful to Martin Guyote for valuable suggestions and comments. Preparation of this article was supported by ONR Contract N0001478C0025 to Robert Sternberg; the research reported in the article was supported by NSF Grant BNS-76-05311 to Robert Sternberg. Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Box 11A Yale Station, Yale University, New Haven, Connecticut 06520.

[1] These six models exhausted the plausible alternatives that we could find that were suitable for our purposes. Since preparation of the article, Johnson-Laird and Steedman (1978) have proposed a new model of syllogistic reasoning. The model of Ceraso and Provitera (1971) was considered, but was deemed unsuitable because of its fundamental assumption that subjects do not know the logical meaning of <u>some</u>. We, like most investigators, instructed subjects as to this meaning. We attempted to quantify the model of Revlis (1975), but found the level of detail presented in Revlis's (1975) article insufficient for quantification.

[2] Root-mean-square deviations were also computed for individual syllogisms. The ideal-subject model had the lowest RMSD or was tied for the lowest on 3 of the 15 syllogisms, the transitive-chain model on 12, the complete-combination model on 5, the random-combination model on 6, the conversion model on 8, and the atmosphere model on 3. These numbers sum to greater than 15 because of the large number of ties.

[3] This observation was made by Martin Guyote.

[4]Three alternative procedures were used for prediction of response choices. In the first, the expected value of the response for each conclusion (on the 1 to 3 scale) was used as the predictor, and the observed response to each conclusion (also on the 1 to 3 scale) was used as the criterion. In the second procedure, computations were identical, except that conclusions designated as logically valid (definitely true) were assigned a value of 1, and conclusions designated as logically invalid (possibly true or never true) were assigned a value of 2. In the third procedure, the expected proportions of definitely trues, possibly trues, and never trues for each conclusion were used as predictors (resulting in three times as many data points as for the first procedure), and the observed proportions were used as criteria. The three methods gave very similar results, and so only results from the first method are reported here. Prediction equations for the method may be obtained by writing to the senior author.

Table 1

Performance of Alternative Models in Predicting Performance on Encoding Task

Group	Conclusions	Number of Data Points	Reliability	Ideal Subject R^2	Ideal Subject RMSD	Transitive-Chain R^2	Transitive-Chain RMSD	Complete Combination R^2	Complete Combination RMSD	Random Combination R^2	Random Combination RMSD	Conversion R^2	Conversion RMSD	Atmosphere R^2	Atmosphere RMSD
Guyote-Sternberg Parameter Estimates															
Subset	Subset	16	.99	.96	.19	.96	.19	.86	.37	.82	.42	.96	.19	.73	.41
Full Set	Subset	16	.99	.97	.17	.97	.17	.87	.35	.83	.41	.97	.17	.74	.41
Full Set	Full Set	40	.99	.95	.16	.95	.16	.81	.37	.74	.45	.80	.37	.33	.67
Number of Parameters				0		0		9		9		0		0	
Sternberg-Turner Parameter Estimates															
Subset	Subset	16	.99	.96	.19	.96	.19	.97	.18	.97	.18	.96	.19	.73	.41
Full Set	Subset	16	.99	.97	.17	.97	.17	.98	.15	.98	.15	.97	.17	.74	.41
Full Set	Full Set	40	.99	.95	.16	.95	.16	.93	.20	.93	.20	.80	.37	.33	.67
Number of Parameters				0		0		9		9		0		0	

Notes: Ideal-subject and transitive-chain models make the same predictions for the encoding task. Parameters were estimated under the constraint that they be nonnegative.

Table 2

Performance of Alternative Models on Encoding Task for Individual Subjects:

Subset Group

Subject Number	Reliability Index[a]	Ideal Subject[b]		Transitive-Chain[b]		Complete Combination[c]		Random Combination[c]		Conversion[b]		Atmosphere	
		R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD
1	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
2	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
3	.90	.84	.35	.84	.35	.79	.43	.76	.47	.84	.35	.51	.61
4	.95	.92	.25	.92	.25	.79	.43	.75	.49	.92	.25	.73	.43
5	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
6	.94	.92	.25	.92	.25	.79	.43	.75	.49	.92	.25	.56	.56
7	.85	.59	.66	.59	.66	.67	.63	.67	.64	.59	.66	.70	.56
8	.98	.92	.25	.92	.25	.79	.43	.75	.49	.92	.25	.73	.43
9	.96	.85	.35	.85	.35	.73	.50	.69	.55	.85	.35	.82	.35
10	.58	.25	.87	.25	.87	.24	.91	.23	.94	.25	.87	.26	.79
11	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
12	.88	.76	.43	.76	.43	.72	.50	.70	.53	.76	.43	.56	.56
13	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
14	.87	.64	.56	.64	.56	.65	.58	.63	.60	.64	.56	.62	.56
15	.98	1.00	.00	1.00	.00	.86	.36	.81	.42	1.00	.00	.67	.50
16	.87	.67	.50	.67	.50	.57	.61	.54	.65	.67	.50	.56	.50
Mean	.92	.84	.28	.84	.28	.74	.47	.71	.52	.84	.28	.63	.52
Best[d]		13	14	13	14	1	0	0	0	13	14	2	5

Table 2 Continued

Note: Parameter estimates used to fit individual data were the group estimates from Experiment 1 of Guyote & Sternberg (Note 1).

[a]Since there were no within-subject replications, the reliability index (square root of the reliability coefficient) of individual data was estimated as the correlation between each subject's individual pattern of data and the group pattern of data (with that individual excluded).

[b]The ideal-subject and transitive-chain models make identical predictions for the encoding task. The conversion model makes predictions identical to these for the subset of conclusions (but not for the full set).

[c]Model fits for the complete- and random-combination models differed slightly because parameters estimated for these models in the Guyote-Sternberg experiment differed slightly.

[d]"Best" refers to the number of individuals for whom the model provided the best fit (including ties).

Table 3

Performance of Alternative Models in Predicting Performance on Combination Task

Group	Conclusions	Number of Data Points	Reliability	Ideal Subject		Transitive-Chain		Complete Combination		Random Combination		Conversion		Atmosphe	
				R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD
Guyote-Sternberg Parameter Estimates															
Subset	Subset	60	.98	.78	.33	.96	.16	.89	.26	.84	.32	.92	.26	.53	.51
Full Set	Subset	60	.98	.88	.24	.94	.20	.87	.28	.81	.35	.84	.35	.46	.55
Full Set	Full Set	150	.97	.82	.27	.87	.26	.85	.28	.81	.34	.71	.42	.20	.75
Number of Parameters				0		4		9		9		0		0	
Sternberg-Turner Parameter Estimates															
Subset	Subset	60	.98	.78	.33	.96	.16	.97	.14	.96	.16	.92	.26	.53	.51
Full Set	Subset	60	.98	.88	.24	.95	.16	.96	.16	.96	.16	.84	.35	.46	.55
Full Set	Full Set	150	.97	.82	.27	.91	.19	.92	.19	.91	.21	.71	.42	.20	.75
Number of Parameters				0		4		9		9		0		0	

Note: Parameters were estimated under the constraint that they be nonnegative.

Table 4

Performance of Alternative Models on Combination Task for Individual Subjects: Subset Group

Subject Number	Reliability Index[a]	Ideal Subject R^2	Ideal Subject RMSD	Transitive-Chain R^2	Transitive-Chain RMSD	Complete Combination R^2	Complete Combination RMSD	Random Combination R^2	Random Combination RMSD	Conversion R^2	Conversion RMSD	Atmosphere R^2	Atmosphere RMSD
1	.88	.48	.66	.74	.46	.72	.48	.69	.51	.72	.48	.45	.68
2	.84	.74	.37	.85	.40	.67	.44	.65	.47	.65	.48	.32	.66
3	.92	.73	.41	.84	.32	.75	.41	.74	.42	.76	.41	.42	.63
4	.76	.46	.61	.59	.53	.62	.52	.59	.55	.56	.58	.28	.73
5	.80	.41	.62	.64	.49	.51	.58	.51	.59	.65	.50	.60	.50
6	.79	.42	.71	.60	.60	.57	.63	.51	.67	.64	.58	.54	.63
7	.85	.53	.56	.71	.44	.60	.53	.55	.57	.74	.43	.41	.65
8	.88	.64	.45	.77	.37	.73	.41	.67	.46	.76	.41	.37	.63
9	.88	.90	.22	.81	.32	.78	.36	.75	.39	.72	.43	.26	.70
10	.86	.69	.43	.74	.40	.72	.42	.67	.47	.69	.47	.34	.67
11	.98	.97	.13	.84	.30	.80	.34	.75	.39	.73	.43	.33	.65
12	.73	.34	.70	.52	.60	.49	.62	.48	.64	.48	.65	.39	.67
13	.89	.58	.58	.87	.34	.67	.51	.54	.62	.83	.37	.51	.63
14	.85	.59	.53	.75	.42	.65	.50	.56	.58	.75	.43	.43	.65
15	.80	.48	.66	.65	.54	.67	.53	.63	.56	.62	.58	.38	.73
16	.77	.56	.52	.60	.50	.63	.49	.65	.48	.54	.58	.22	.75
Mean	.84	.60	.51	.72	.44	.66	.49	.62	.52	.68	.49	.39	.66
t^b		2	3	7	8	2	2	1	1	4	2	0	0

Table 4 Continued

Note: Parameter estimates used to fit individual data were the group estimates from Experiment 1 of Guyote & Sternberg (Note 1).

[a]Since there were no within-subject replications, the reliability index (square root of the reliability coefficient) of individual data was estimated as the correlation between each subject's individual pattern of data and the group pattern of data (with that individual excluded).

[b]"Best" refers to the number of individuals for whom the model provided the best fit (including ties).

Table 5

Parameter Estimates for Alternative Models

Parameter	Transitive-Chain[a]		Complete Combination[b]			Random Combination[b]		
Data Set........	G-S	S-T	G-S	S-T	S-T	G-S	S-T	S-T
Task[d]..........	Comb.	Comb.	Comb.	Comb.	Enc.	Comb.	Comb.	Enc.
p_1	.54	.54	.68	.80	.69	.62	.80	.69
p_2	.13	.02	.32	.19	.32	.38	.19	.32
p_3	.00	.24	.00	.00	.00	.00	.01	.00
p_4	.33	.19	.07	.44	.44	.00	.44	.44
p_5			.19	.29	.00	.05	.38	.00
p_6			.74	.27	.53	.95	.17	.53
p_7			.12	.08	.00	.06	.16	.00
p_8			.64	.55	.53	.76	.42	.53
p_9			.24	.39	.47	.18	.43	.47

Note. Complete and random combination models make indistinguishable predictions for the encoding task.

Table 5 Continued

[a]p_1 = P(Performing exactly one combination)

p_2 = P(Performing exactly two combinations)

p_3 = P(Performing exactly three combinations)

p_4 = P(Performing exactly four combinations)

[b]p_1 = P(encoding "All A are B" as A equivalent to B)

p_2 = P(encoding "All A are B" as A subset of B)

p_3 = P(encoding "Some A are B" as A superset of B)

p_4 = P(encoding "Some A are B" as A overlapping with B)

p_5 = P(encoding "Some A are B" as A equivalent to B)

p_6 = P(encoding "Some A are B" as A subset of B)

p_7 = P(encoding "Some A are not B" as A superset of B)

p_8 = P(encoding "Some A are not B" as A overlapping with B)

p_9 = P(encoding "Some A are not B" as A disjoint with B)

[c]G-S = Guyote-Sternberg data

S-T = Sternberg-Turner data

[d]Comb. = Combination task

Enc. = Encoding task

Table 6

Example of a Modified Truth-Table Analysis

Premises: All B are C.
All A are B.

Conclusion	Set Relation(s) Used to Represent A-C Relation		
	A equivalent to C	A subset of C	A equivalent to C or A subset of C
All A are C.	D	D	D
No A are C.	N	N	N
Some A are C.	D	D	D
Some A are not C.	N	N	N
All C are A.	D	N	P
Some C are not A.	N	D	P
All A are C and all C are A.	D	N	P
Some but not all A are C and some but not all C are A	N	N	N
All A are C and some C are not A	N	D	P
All C are A and some A are not C	N	N	N

Note: D = Definitely, P = Possibly, N = Never.

Table 7

Representations of Encoded and Combined Premises

Encoding Task

Set Relation

Premise	A,B coincident	A,B overlapping	A subset of B	A superset of B	A disjoint with C
All A are B.	x		x		
No A are B.					x
Some A are B.	x	x	x	x	
Some A are not B.		x		x	x

Combination Task

Set Relation

Premises	A,C coincident	A,C overlapping	A subset of C	A superset of C	A disjoint with C
All B are C. All A are B.	x		x		
All C are B. No A are B.					x
No B are C. All A are B.					x
All C are B. Some A are not B.		x		?[a]	x
No C are B. Some B are A.		x			x

Table 7 Continued

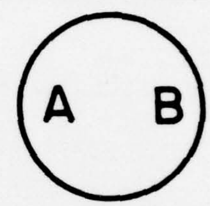
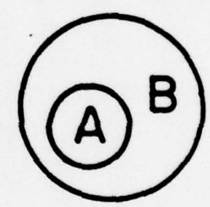
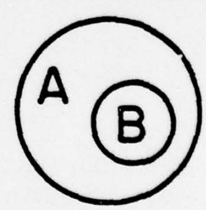
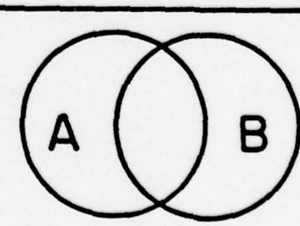
Premises	A,C coincident	A,C overlapping	A subset of C	A superset of C	A disjoint with C
Some B are not C. All B are A.		x		x	x
No C are B. All B are A.		x		x	x
All C are B. All B are A.	x			x	
All B are C. All B are A.	x	x	x	x	
Some B are C. All B are A.	x	x	x	x	
All B are C. Some A are B.	x	x	x	x	
All B are C. No A are B.		x			x
Some C are not B. All A are B.		x	x		x
Some C are B. Some B are not A.	x	x	x	x	x
Some B are not C. No B are A.	?[a]	x	x	x	x

Note: An x in a given cell indicates that this set relation is part of the encoded or combined representation.

[a]Due to a tie between RMSDs for two representations, the inclusion of this set relation is questionable.

Figure Caption

Figure 1. Representations of set relations expressed as Euler diagrams.

Set Relation	Euler Diagram Representation
Equivalence	
Subset-Set	
Set-Subset	
Overlap	
Disjoint